利用稀疏语义结合双层深度卷积神经网络的敏感图像检测方法 *

如先姑力·阿布都热西提 [1]，亚森·艾则孜 [1†]，孙国梓 [2]

(1. 新疆警察学院 信息安全工程系，乌鲁木齐 830013; 2. 南京邮电大学 计算机技术研究所，南京 210003)

摘　要：互联网技术的飞速发展导致敏感内容图像由原先基本隐蔽的内容交换变为海量的数据共享，传统基于图像特征提取的敏感内容检测方法不再适用。针对上述难点，提出基于稀疏语义和双层深度卷积神经网络相结合的敏感内容检测方法。上层网络首先进行训练样本的预处理，并通过构造图像的稀疏语义表示作为神经网络的输入，而下层网络则进一步考虑第三方管控机制（如政府代理等），提出针对特定群体的敏感内容图像检测方法。与现有常用敏感内容图像检测方法相比，提出的检测方法可有效降低训练样本数量，且检测精度比传统图像检测方法（如基于视觉词袋方法等）提升 7%以上。

关键词：敏感图像内容检测；双层卷积神经网络；深度学习算法；稀疏语义表示；视觉词袋；皮肤检测器

中图分类号：TP391.41　　doi: 10.19734/j.issn.1001-3695.2018.11.0851

Sensitive image detection method based on deep neural network and sparse semantics

Ruxianguli·Abudurexiti[1], Yasen·Aizezi[1†], Sun Guozi [2]

(1. *Dept. of Information Security Engineering*, *Xinjiang Police College*, *Urumqi* 830011, *China*; 2. *Institute of Computer Technology*, *Nanjing University of Posts & Telecommunications*, *Nanjing* 210003, *China*)

Abstract: With the rapid development of Internet technology, sensitive content images have changed from basic concealed content exchange to mass data sharing. The traditional method of sensitive content detection based on image feature extraction is no longer applicable. To overcome these difficulties, this paper proposes a sensitive content detection method based on sparse semantics and double-layer deep convolution neural network. In this method, the upper network preprocesses the training samples and constructs sparse semantic representation of the image as the input of the neural network, while the lower network further considers the third-party control mechanism (such as government agents) and proposes a sensitive content image detection method for specific groups. Compared with the existing image detection methods for sensitive content, the proposed method can effectively reduce the number of training samples, and the detection accuracy is more than 7% higher than that of traditional image detection methods (such as visual word bag method) .

Key words: sensitive image detection; double-layer artificial neural network; deep learning algorithm; sparse semantic representation; visual word bag; skin detector.

0　引言

近年来，随着社交互联网络的兴起与发展，新兴的通信和计算机技术的进步极大地促进了世界范围内媒体数据的传播。然而，这些技术的发展也使得人类能够更方便地接触到具有敏感内容的图像信息。尤其对于特定人群而言，这些敏感内容的图像将对其成长发展造成严重的、不可逆的误导。自 20 世纪 90 年代中期以来，敏感内容图像在生成、传播与存储等环节已发生了巨大的变化，从原先基本隐蔽的内容交换演变为大规模的海量数据共享。因此，针对互联网中存在的海量敏感图像内容进行准确甄别和检测具有十分重大的研究意义。

近年来，国内外众多科研人员提出了敏感图像内容检测算法。例如，文献[1]利用深度可分离卷积神经网络和 MobileNet 模型，配合 GPU 并行计算架构建立了对敏感图像具有较高准确甄别度的识别模型,但所提方法难以有效甄别看似无害、隐蔽程度高的敏感图像。文献[2]提出一种以数据为导向的基于深度卷积神经网络（convolution neural network, CNN）来获取图像特征的敏感图像检测方法，使用批量随机梯度下降算法训练卷积神经网络以获取图像特征,但批量梯度下降算法的训练精度与训练速度间的折中选取依赖于人为经验设定，且难以保证获取全局最优解。文献[3]提出一种利用基于支持向量机(support vector machine, SVM)的敏感图像快速过滤算法，采用混合肤色模型实现裸露皮肤区域快速检测，从而提取人脸位置、形状和图像背景等特征,但 SVM 算法难以适用于样本数量较多情形，即对计算机内存与运算时间提出较高要求。类似关于敏感内容图像检测的相关工作参见文献[4~7]。此外，上述文献并未考虑政府管控背景下的敏感内容检测技术，即无法对敏感内容图像进行人为的干预检测。

针对上述问题，本文提出一种基于稀疏语义的双层深度卷积神经网络敏感内容图像检测方法。其中，上层采用一般性的训练样本对人工神经网络进行训练，实现非敏感、特定人群以及疑似无害三类图像的初步分类；下层通过引入适当

收稿日期：2018-11-22；**修回日期：**2018-12-29　**基金项目：**国家自然科学基金资助项目（61762086）；新疆维吾尔自治区高校科研计划面上项目（XJEDU2017M046）

作者简介：如先姑力·阿布都热西提（1976-），女（维吾尔族），新疆喀什人，副教授，硕士，主要研究方向为信息安全、自然语言处理等；亚森·艾则孜（1975-），男（维吾尔族）（通信作者），新疆库车人，国家电子数据司法鉴定员，教授，硕士，主要研究方向为数字取证、自然语言处理等(yasenaizezi@126.com)；孙国梓（1972-），男，安徽天长人，教授，博士，主要研究方向为网络空间安全.

的政府代理机制实现专门的针对特定人群敏感内容图像检测的适当调整。实际算例表明，本文提出的敏感内容图像检测方法相比现有敏感内容图像检测方案可更加有效地减少训练样本数量，同时降低检测误差。而所提出的人工干预机制，可实现政府代理人（如执法机构，专业部门等）的灵活控制。

1　图像的稀疏语义表示

为便于后续神经网络的训练以及最终将图像分类为敏感内容图像以及非敏感内容图像，本文首先将图像特征进行稀疏化表示。对于包含 k 个图像特征的集合

$$F=\{f_1, f_2, \dots, f_k\} \tag{1}$$

假定由第 i (i=1，2)类训练图像生成的码本为

$$A_i=[w_{i,1}, w_{i,2}, \dots, w_{i,C_1}]\in\Omega^{128*C_1} \tag{2}$$

其中：C_1 为第 f 个图像特征类码本的单词数目，对于样本图像 $y_0\in\Omega^{128*C_1}$，其可表示为 A_i 中码本的线性近似表示，即

$$y_0\approx\sum_{j=1}^{C_1}a_{i,j}w_{i,j}=A_i a_i \tag{3}$$

其中：$a_i=[a_{i,1}, a_{i,2}, \dots, a_{i,C_1}]^{\mathrm{T}}\in\Omega^{C_1}$。更一般地，对于 K 类图像，则有 $A=[A_1, A_2, \dots, A_K]$，而 y_0 线性表示为

$$y_0=Aa \tag{4}$$

其中：$a=[a_1, a_2, \dots, a_K]$，a 是 C_1 维的稀疏解向量，为满足下式的最优解：

$$\hat{a}=\arg\min\|y-Aa\|_2 \tag{5}$$

其中：L_0 范数$\|.\|_2$ 用于度量解向量 a 中非零元素的个数。当解向量 a 足够稀疏时，采用特征码本的最优近似线性表示即为求解 a 的最小 l_1 范数。因此可以转换为如下优化问题：

$$\begin{cases}\mathbf{OP}: & \min\|a\|_1\\ \mathbf{s.t.}: & \hat{a}=\arg\min\|y-Aa\|_2^2\end{cases} \tag{6}$$

根据拉格朗日定理，优化问题式(6)可进一步通过优化求解如下的目标函数：

$$E(a;\ \lambda)=\|y-Aa\|_2^2+\lambda\|a\|_1 \tag{7}$$

其中，参数 λ>0 是逼近项和稀疏项之间的平衡因子。

基于稀疏语义表示的物理意义为，当向量 a 的非零系数值反映图像样本与构建的图像特征码本的相关性。相关性越强，采用该类码本对图像特征进行描述效果越好，即重构误差式(8)越小；反之则钩虫误差越大。

$$\varepsilon_1=\sum_k\|y-y_0\|_2^2 \tag{8}$$

2　利用深度 CNN 的敏感内容图像检测方法

2.1　问题分析

传统图像识别算法通过提取颜色、形状、纹理以及其他人工描述项的图像特征来区分图像之间的细微差别，但应用此类方法筛选无害和非敏感内容是极其困难的。一种简便的方法是通过检测因特网下载的数据量大小与相关服务进行标记，这也是近几年深度学习理论在计算机视觉领域研究大放异彩的重要原因。然而针对特定敏感群体（如敏感人）的敏感内容图像的识别与检测，不仅需要关注图像的检测分类，还需要对图像内容进行更严格的控制与甄别，这也是目前面向特定群体的基于深度学习敏感内容图像检测难点所在。

针对上述实际问题提出如下解决思路：基于深度学习基本算法架构，但摈弃从零开始训练数据样本的传统方式，神经网络中各神经元节点的权重系数采用来源于源问题的侧面数据进行初始化，随后使用训练后（即权重系数调整后）的样本回代至目标求解问题，从而提升敏感内容图像间不同细微差别的检测精度。遵循该思路，本文提出一种基于深度卷积神经网络的分层图像检测方案。其中，上层为权重系数确定环节，即利用深度卷积神经网络对一般图像进行分类检测，初步确定神经网络各神经元节点权重；下层引入政府代理人控制机制，针对特定人群敏感内容图像检测问题人为地对权重系数进行微调，并再度进行神经网络的训练。本文所提方法的一个显而易见的优点为：由于上层伸进网络已专门针对网络权重进行了初始化故，下层所需训练样本数量更为精简。

2.2　双层深度卷积神经网络架构

提出的双层深度 CNN 基本架构如图 1 所示。其中，第一层网络为源问题求解网络，其基本目标（即源任务）为对图像进行合理的分类。一般来说，图像检测源任务需将图像分类为涵盖家宠、家庭用品、食物、交通工具等在内的 1000 中日常类别中。本文所提算法结构包含一个初始化模块，9 个处理模块，2 个辅助分类器和一个最终分类器。图 2 则显示了每个模块的技术细节。各模块的主要功能解释如下：

初始化模块主要通过卷积、池化、局部响应归一化等操作对输入的原始图像进行处理，本文中采用 7×7 卷积内核对原始图像进行解析，而后逐步采用卷积、规范化和数据池化操作进行数据量化加工。图像初始化操作完成后得到的数据通过 7 个核心的特征学习模块进行学习与训练，每个模块都可看成一个小型神经网络，可实现图像分类源任务中特定图像特征与图像属性的检测。

除了 7 个特征学习模块之外，本文另外在第 3 和第 6 个特征学习模块后设置了两个辅助分类器以便通过网络放大梯度信号，实现早期特征数据的突出显示。与特征学习模块类似，这两个辅助分类器同样也是一个小的卷积网络。此外，本文中辅助分类器包括一个 softmax 分类决策层和一个数据缩减层，通过平均池与 5×5 内核。需要说明的是，辅助分类器仅在训练期间使用，而在测试期间则会移除。

最终分类器的基本功能是分类。在一般性的源问题中，最终分类器包含一个数据缩减层，通过内核的平均池、全连接化和具有 softmax 损失的线性层，实现 1000 种图像类型的辨识。然而考虑到本文中的两个目标问题（即非敏感内容和特定人群敏感内容图像检测），本文将最终分类器采用径向基函数（radial basis function，RBF）内核的双向非线性 SVM 进行设计，其原因在于，当执行针对特定群体的敏感内容图像检测时，使用该分类器具有更高的检测辨识率。

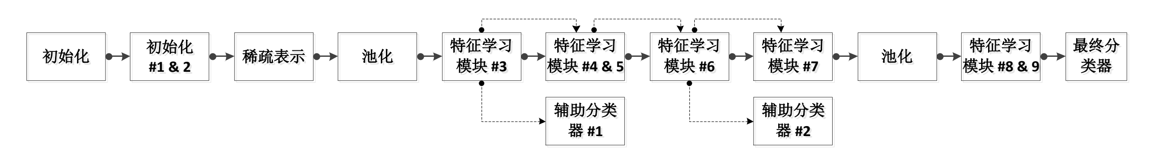


图 1　对象分类源问题的所选架构概述

Fig. 1　Overview of selected architecture for object classification problem

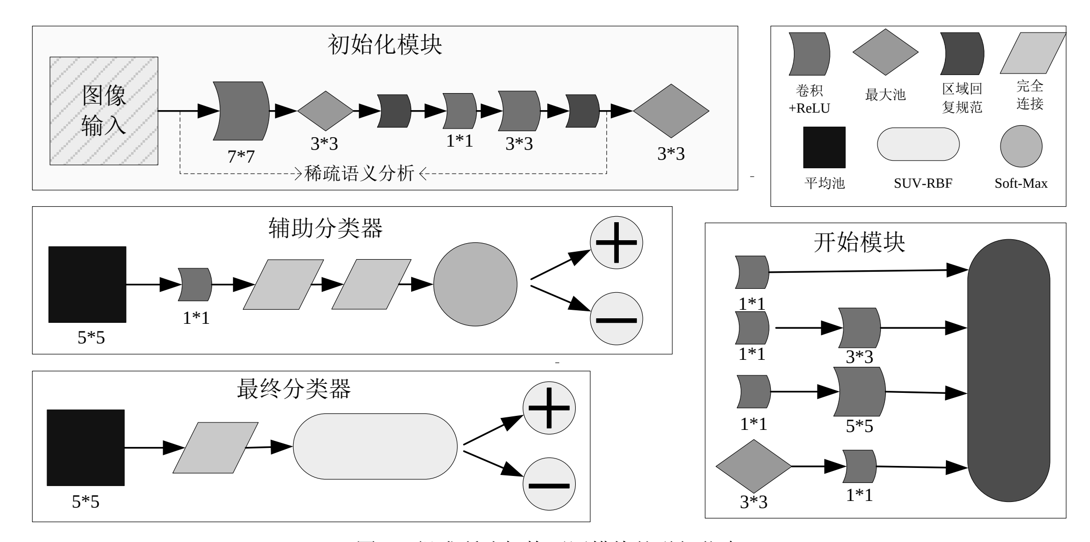


图 2　组成所选架构不同模块的详细信息

Fig. 2　Details of the different modules of the selected architecture

2.3　上层源问题网络训练

针对源问题的神经网络训练样本大约需要 120 万幅包含 1000 种不同种类的图像。此过程中采用带动量（设为 0.9）的异步随机梯度下降优化算法以及固定学习速率（每隔 8 个学习周期使学习速率下降 4%）。源问题网络训练阶段，使用两个辅助分类器对每幅图像的观测信息损失量乘以权重系数 0.3 后附加到最终分类器的观测信息损失量中。为收集更多的训练样本，采用旋转、镜像和裁剪但不包括缩放、光度畸变等操作对初始池进行动态扩充。此外，本文中采用多项式学习速率衰减策略来代替阶跃变化的学习速率衰减策略，从而达到 4 倍甚至更快的训练速度。

上层神经网络的训练步骤为：对于一组给定的训练输入样本，所有的图像均被首先采用保持纵横比的方式重新调整大小，且所有输入样本图像的大小大于本文所设定的输入阈值（224×224 像素）；随后，图像被进一步裁剪至特定卷积网络所需的形状。值得一提的是，对输入样本图像的预处理过程无论是对于源问题实现图像的基本分类以及针对特定群体的图像检测分类都是完全相同的。

2.4　迁移学习与针对目标问题的调整策略

当经过训练的神经网络可对 1000 种目标类别中的日常图像进行准确分类后，本文进一步提出针对目标检测问题的学习权重进行微调，其目的是通过使用相同的网络架构进行针对特定群体的敏感图像检测。本文中下层目标本文求解网络的神经元权重并非从头开始训练来得到，而是采用两路 softmax 分类器（即输出非敏感图像或特定人群敏感内容），并对初始权重进行微调，并将标记为非敏感或特定人群敏感的系列图像作为输入进行反向传播。此外，两个辅助分类器中同样采用两路 softmax 分类函数。

当重新调整权重后的网络能够准确进行非敏感内容图像的检测时（即网络收敛时），最后的分类函数采用带 RBF 内核的 SVM 分类器进行代替。此项操作使得深度学习神经网络具有专门针对非敏感内容图像检测的权重，并可使神经网络工作于非敏感内容图像分类的特征提取器，而 SVM 分类器则在这些特征之上进行特定问题求解模型的学习。由于本文的最终目标值对特定人群敏感图像进行检测，故可将非敏感内容图像检测作为神经网络工作可靠性的测试实例，通过输入特定人群敏感图像和非特定人群敏感图像实例，进行图像特征量的提取以及特定人群敏感图像检测的 SVM 分类器训练。具体训练步骤如下：

a)给定一个包含特定人群敏感内容和非特定人群敏感内容的图像集合，使用训练后的神经网络提取其图像特征并使用这些示例图像进行非线性 SVM 分类器的训练。为训练 SVM 分类器，本文在可用的训练集中进行 5 次交叉验证策略，并采用网格化搜索方式以寻找最佳分类参数 $C\in\{2^c\text{: } c\in[-5, -3, -1,\ldots, 15]\}$）以及 $\gamma\in\{2^i\text{: } i\in[15, 13, \ldots, 3]\}$。最终获得的第一组 SVM 分类器参数为 γ=0.0078125 以及 C=8.0。该检测网络可称为初级检测器，可检测非敏感内容图像，还在某种程度上检测特定人群敏感内容图像，但此时的神经网络仍不具备最终的针对敏感群体的敏感内容图像的检测内容。

b)在将初始针对图片分类的神经网络调整为敏感内容分类检测的神经网络后，本文重点关注将神经网络进一步微调至具备针对敏感群体的敏感内容图像检测,为此提出两种解决途径。对于第一种解决途径，从受过对象分类训练的网络开始，对其学习的权重进行微调，接收一系列标记为敏感图像和非敏感图像内容，并将使用的图像作为输入反向传播。与前述内容一致，当网络训练收敛得到的权重可实现针对目标问题（即特定人群敏感内容图像检测）的检测是，将最后一层的分类器替换为具有 RBF 内核的 SVM 分类器。本文称此时的第二层神经网络为第一类特定人群敏感内容图像检测器（称为方案 1）。这是由于此时针对目标的特定人群敏感内容图像检测是通过从源问题直接调整过来的（图像种类分类→特定人群敏感内容图像检测）。

对于第二种解决方案，本文从网络微调到非敏感内容检测问题开始，并对其权重进行微调，接收一系列标记为 SEIC 和非 SEIC 内容的图像，并再次使用反向传播。与前述内容一致，当网络训练收敛得到的权重可实现针对目标问题（即特定人群敏感内容图像检测）的检测是，将最后一层的分类器替换为具有 RBF 内核的 SVM 分类器。本文称此时的第二层神经网络为第二类特定人群敏感内容图像检测器（称为方案 2）。这是由于本文使用了两个步骤进行权重的调整（图像种类分类→非敏感内容检测→特定人群敏感内容图像检测）。

对于给定的针对特定人群敏感内容图像或非特定人群敏感内容图像集合，本文使用方案 1 或方案 2 中的训练后的神经网络进行图像特征的提取，并采用这些图像实例进行非线性 SVM 分类器的训练。对于训练过程，本文同样采用可用的数据集进行 5 次交叉验证，并采用网格化搜索方式以寻找最佳分类参数 $C\in\{2^c\text{: } c\in[-5, -3, -1,\ldots, 15]\}$）以及 $\gamma\in\{2^i\text{: } i\in[15, 13, \ldots, 3]\}$。最终计算得到的第一组 SVM 分类器参数数

为 γ=0.0078125 以及 C=0.5。

2.5 样本数据的扩充

即便在对神经网络权重进行微调而不是重新开始训练，同样需要大量的样本数据。因此，针对本文所提双层人工神经网络，本文提出如下解决方案以实现在权重系数重新调整时的样本数据扩充。其中，由于用于微调的数据集很少包含实际负面的例子，故而初始训练中进一步增强了非敏感特定人群内容具体是指非敏感图片。因此，与从原始图像生成额外图像（如旋转、镜像或是调整对比度等）不同，本文所使用的数据集采用来自 Microsoft 公共对象数据集中的 20000 个图像，其中包含 16000 个共计 91 个常见类别。对于第一层源问题求解，本文不进行数据扩充；而对于第二层目标问题的求解，本文进行有样本数据扩充以及无样本数据扩充的两组实验。

3 图像分类实验参数设置

3.1 数据集

为验证本文所提图像分类检测算法的有效性，本文选用 Pornography-2k 数据集[8]为实验对象进行分析。其中包括近 140 h 的 1000 个含敏感内容和 1000 个不含敏感内容的视频，视频长度从 6 s 到 33 min 不等。本文进一步将原视频分解为图像帧，从而得到共计 58 971 张图像，其中 33 723 张图像包含敏感图像内容，即敏感图像和非敏感图像内容的比例分别为 57%和 43%。图 3 示出了所选数据集的部分示例帧。共计依据上节所述算法步骤，本文中将数据集分为训练集（含 33 646 个图像）、验证集（含 5 938 个图像）和测试集（含 19 387 个图像）三种类型。

对于敏感内容图像的分类检测，首先使用训练集对卷积神经网络进行预训练，从而得到卷积神经网络中的各项关键参数。随后使用验证数据集对上述参数进行优化调整。最后利用测试集进行算法有效性验证与分析。其中，本文采用标准的随机 5×2 交叉验证机制，通过进行随机拆分数据集的方式进行算法有效性的评估。

3.2 分类指标

采用 Pornography-2 k 数据集的默认图像分类度量值，即归一化度量 ACC 和 F_2 度量。其中，ACC 表征正确分类图像的百分比，而 F_2 度量则指的是准确率（precision）与召回率（recall）间的加权调和平均值，如式（9）所示，其中 β=2。

$$F_\beta=(1+\beta^2)\times\frac{\text{precision}\times\text{recall}}{\beta^2\times\text{precision}+\text{recall}} \tag{9}$$



图 3 Pornography-2 k 数据集的示例帧

Fig. 3 Example frame of Pornography-2 k dataset

3.3 取证工具

为了评估所提方案的分类性能，选择依赖于可视内容的取证工具包括 NuDetective[9]、MediaDetective[10]和 Snitch Plus[11]。NuDetective 工具是一种随时可用于研究目的的工具，可用于检测敏感图像内容，而 MediaDetective 工具和 Snitch Plus 工具都属于商业解决方案，其可使检测图像敏感内容相对集中，以上所有工具均以皮肤检测器为基础从而识别敏感内容图像。

此外，对于 MediaDetective 工具和 Snitch Plus 工具，图像文件会根据其对敏感图像内容的可疑程度（即概率）进行评级。如果图像的返回概率等于或大于 50%，会将该图像标记为敏感图像。而 NuDetective 工具则会为图像分配二进制标签：正（即图像是敏感图像）或负（即图像是非敏感图像）。

MediaDetective 工具和 Snitch Plus 工具有四种预定义的执行模式，主要取决于皮肤探测器的严谨程度，在本次实验中，将会选择最严格的执行模式，同时对 NuDetective 工具采用默认设置。

3.4 皮肤检测器

当前，人们已经提出了多种方法来检测敏感图像内容，但人类皮肤检测技术[12]，尤其对于基于颜色的检测技术，在识别的简单性和准确性的提高上仍然具备挑战性。现有方法主要通过定义 RGB 颜色空间中的皮肤区域，从而提出皮肤分类。皮肤分类模型可大致分为遵循以下三个准则：

准则 1：

$$[(R>95)\wedge(G>40)\wedge(B>20)]\wedge \tag{10}$$

准则 2：

$$[\max(R,G,B)-\min(R,G,B)>15]\wedge \tag{11}$$

准则 3：

$$[|R-G|>15\wedge(R>G)\wedge(R>B)] \tag{12}$$

其中：R、G 和 B 表示 RGB 颜色空间中的像素值，其值范围为 0~255。为方便比较，本文选择了一种被广泛使用并且易于实施的皮肤检测器，如果图像中的皮肤像素百分比等于或大于 13%，则将图像分类为敏感图像，以此在训练集中执行交叉验证，以选择最佳阈值。

3.5 视觉词袋

在引入深度学习技术之前，视觉词袋（bag of visual words，BoVW）[13]建模方法是最常用的描述图像内容的方法，它将图像表征为通过量化局部描述符的空间。

本文对两种经典的 BOVW 方法（即硬编码方法和平均池化方法）进行评估。首先将图像样本进行预处理，得到最大不超过 10 万像素的数据集，随后采用步长分别为 4、6、8、11 和 16 像素生成的 24×24、32×32、48×48、68×68 和 96×96 像素网格进行图像局部特征量（本文使用加速稳健特征（speeded-up robust features, SURF）描述符）的提取，并采用主成分分析（principal component analysis）方法将 SURF 的维数降低到 32 维。

视觉词库通过采用基于欧几里德距离定义的 K-means 聚类工具，从超过 100 万个随机抽样的 PCA 描述符中进行学习，并从中提取出 2 048 个视觉词，而分类器则采用 SVM 分类器。默认情况下，本文使用带 RBF 内核的非线性 SVM 分类器，

并采用网格化搜索方式以寻找最佳分类参数 $C\in\{2^c: c\in[-5, -3, -1,\ldots, 15]\}$）以及 $\gamma\in\{2^i: i\in[15, 13, \ldots, 3]\}$。最终计算得到的 SVM 分类器参数为 γ=0.0078125 以及 C=32。

4　实验与结果讨论

为了验证所提方法在敏感图像检测效果方面的优异性和有效性，将通过设计实验与当前流行的其他检测方法、取证工具和商业软件进行比较，从而根据检测敏感图像内容的实验结果进行分析。本文基于 Maltab 2014a 软件进行所提相关算法的编写，且计算机配置为：操作系统 Windows 10，内存 2 GB，硬盘 256 GB，CPU 型号为 I5-3230M (2.6 GHz，双核)，GPU 型号为 Nvidia GeForce GT 740M (2 GB)。

4.1　非敏感内容检测结果

首先利用 Pornography-2k 数据集（共计 257 522 帧）对不同的非敏感内容图像检测性能进行分析。本文选用文献[14]所提的检测方法与本文所提方法进行对比。表 1 给出了两种方法下的 ACC 度量指标和 F_2 度量指标，可以看出，本文所提方法对图像的检测精度优于文献[14]检测方法。而在计算资源方法，本文所提方法同样更具经济性，文献[14]采用 16 个 GPU 集群进行神经网络的训练，而本文则仅采用单个 GPU。此外，虽然两个网络在实际非敏感内容检测方面具有相似的检测性能，但本文所提方法在拦截敏感内容图像的检测精度比文献[14]高 3.5%。

表 1　敏感内容检测结果

Table 1　Sensitive content detection results　/%

比较方法	TPR	TNR	F_2	ACC
文献[14]方法	94.5	82.3	94.1	88.4
提出的方法	94.8	88.2	94.8	91.5

TPR 表示真正值率；TNR 表示真负值率；ACC 表示精确度；F_2 表示 F_2 测量值

4.2　特定人群敏感图像检测

进一步考虑本文的最终目标，即敏感内容图像检测。为对比说明本文方法的优越性，采用表 2 中所示的常用方法作为对比加以说明。如表 2 所示，本文采用的基于皮肤检测器的系统在性能上相比较与其他系统有了相当大的改善，对特定人群敏感内容的检测精度有了较大提升。

针对非敏感内容的图像分类性能，与 BoVW 和文献[14]中雅虎公司开发算法相比，其 ACC 指标和 F_2 指标分别提升了 7.7%和 6.5%，而针对特定人群敏感内容的图像分类性能，本文所提分类算法较之于文献[14]，其 ACC 指标和 F_2 指标则分别提升了 8.6%和 7.0%。

表 2　敏感图像内容检测结果

Table 2　Results of sensitive image content detection　/%

比较方法	TPR	TNR	F_2	ACC
BoVW[15]	69.9	66.6	71.1	68.3
文献[14]	80.1	74.9	80.7	77.5
非敏感内容分类器	83.0	77.8	83.4	80.4
特定人群敏感内容分类器	87.2	85.0	87.7	86.1

图 4 示出了不同分类器图像检测误差率直方图，本文所提特定人群敏感内容检测方法与基于皮肤检测方法相比，其误差率由 46.9%降低至 3.9%，误差率减少了 70.4%；而与基于 BoVW 的方法相比，误差率则降低了 56.2%，从而表明，基于 BoVW 技术的图像检测精度比皮肤分类器检测精度要高，而本文基于深度卷积神经网络的检测精度则可进一步降低图像误检率。此外，从表 2 和图 4 中也可知，本文所提的针对特定人群敏感内容的检测方法可进一步提升针对非敏感内容检测的神经网络检测精度。

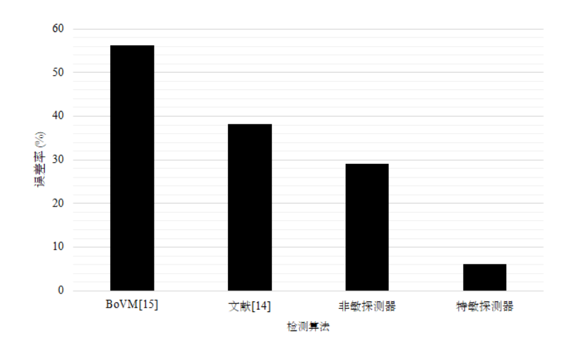


图 4　各探测器误差降低率

Fig. 4　Error reduction rate of detectors

5　结束语

深度卷积神经网络进行图像分类虽具有较高的检测精度，但前期针对神经网络的训练需要大量的原始样本数据制约了其在敏感内容图像中的应用。针对这一问题，本文提出了一个基于稀疏语义与深度卷积神经网络相结合的解决方案，并提出双层神经网络算法进行目标问题求解时神经元权重系数的优化求解。仿真算例结果表明，本文所提方法在检测精度上比常规算法具有较大提升，针对特定人群敏感内容图像的检测性能由于常规基于视觉词袋以及基于皮肤分类的检测方法。

参考文献：

[1] 邢艳芳，卓文鑫，段红秀. 基于 MobileNet 的敏感图像识别系统设计 [J]. 电视技术, 2018, 42(7): 53-56. (Xing Yanfeng, Zhuo Wenxin, Duan Hongxiu. Design of sensitive image recognition system based on MobileNet [J]. Television Technology, 2018, 42(7): 53-56.)

[2] 余明扬，羊鹏，王一军. 基于卷积神经网络的色情图像检测 [J]. 计算机应用与软件, 2018, 35(1): 232-236. (Yu Mingyang, Yang Peng, Wang Yijun. Pornographic image detection based on convolution neural network [J]. Computer Applications and Software, 2018, 35(1): 232-236.)

[3] 周建政，陈法叶，姚金良. 一种基于 SVM 的网络不良图像过滤方法 [J]. 计算机应用与软件, 2012, 29(5): 251-253. (Zhou Jianzheng, Chen Faye, Yao Jinliang. A network bad image filtering method based on SVM [J]. Computer application and software, 2012, 29(5): 251-253.)

[4] 王景中，周靖. 基于比例特征的网络不良图像过滤算法研究 [J]. 计算机工程与科学, 2016, 38(3): 514-519. (Wang Jingzhong, Zhou Jing. Research on network bad image filtering algorithm based on proportional feature [J]. Computer Engineering and Science, 2016, 38(3): 514-519.)

[5] Bissias G, Levine B, Liberatore M, *et al.* Characterization of contact offenders and child exploitation material trafficking on five peer-to-peer networks. [J]. Child Abuse & Neglect, 2015, 52: 185-199.

[6] 刘兴旺，王江晴，徐科. 一种融合 AutoEncoder 与 CNN 的混合算法用于图像特征提取 [J]. 计算机应用研究, 2017, 34(12): 3839-3843. (Liu Xingwang, Wang Jiangqing, Xu Ke. A hybrid algorithm combining AutoEncoder and CNN for image feature extraction [J]. Application Research of Computers, 2017, 34(12): 3839-3843.)

[7] 胡长胜，詹曙，吴从中. 基于深度特征学习的图像超分辨率重建 [J]. 自动化学报, 2017, 43(5): 814-821. (Hu Changsheng, Zhan Shu, Wu Zhongzhong. Image super-resolution reconstruction based on depth feature learning [J]. Journal of Automation, 2017, 43(5): 814-821.)

[8] 贾旭，孙福明，李豪杰，等. 具有普适性的改进非负矩阵分解图像特征提取方法 [J]. 计算机应用，2018,38(1):233-237. (Jia Xu, Sun Fuming, Li Haojie, etc. Universal improved non-negative matrix decomposition image feature extraction method [J]. Computer applications, 2018, 38 (1): 233-237.)

[9] 任晓芳，秦健勇，杨杰，等. 基于能量模型的 LS-TSVM 在人体动作识别中的应用[J]. 计算机应用研究,2016, 33(2): 598-601. (Ren Xiaofang, Qin Jianyong, Yang Jie, *et al.* Application of LS-TSVM based on energy model in human motion recognition [J]. Application Research of Computers, 2016, 33(2): 598-601.)

[10] Perez M, Avila S, Moreira D, *et al.* Video pornography detection through deep learning techniques and motion information [J]. Neurocomputing, 2016, 230 (C): 279-293.

[11] Kanthan S,Graham J A, Azarchi L. Media detectives: bridging the relationship among empathy, laugh tracks, and gender in childhood [J]. Journal of Media Literacy Education, 2016, 8(2): 35-53.

[12] Niezgoda S, Way T P. SNITCH: a software tool for detecting cut and paste plagiarism [J]. ACM SIGCSE Bulletin, 2006, 38 (1): 51-55.

[13] 陈小雪，尉永清，任敏，等. 基于萤火虫优化的加权 K-means 算法 [J]. 计算机应用研究，2018, 35(2): 466-470. (Chen Xiaoxue, Wei Yongqing, Ren Min, *et al.* Weighted K-means algorithm based on firefly optimization [J]. Application Research of Computers, 2018, 35(2): 466-470.)

[14] Mahadeokar J, Pesavento G. Open sourcing a deep learning solution for detecting nsfw images[EB/OL].(2016-09-30). https://noise.getoto.net/2016/09/30/open-sourcing-a-deep-learning-solution-for-detecting-nsfw-images/.

[15] 王茜，陈一民，丁友东.复杂环境中基于视觉词袋模型的车辆再识别算法 [J]. 计算机应用，2018, 38(5): 1299-1303. (Wang Qi, Chen Yimin, Ding Youdong. Vehicle recognition algorithms based on visual word bag model in complex environment [J]. Computer Applications, 2018, 38(5): 1299-1303.)